AD-A107 827 HARVARD UNIV CAMBRIDGE MASS DEPT OF STATISTICS F/G 12/1
ROBUST PRINCIPAL COMPONENTS AND DISPERSION MATRICES VIA PROJECT--ETC(U)
SEP 81 Z CHEN, G LI N00014-79-C-0512
UNCLASSIFIED PJH-8 NL

1 OF 1
AD A
107827

END
DATE
FILMED
1 -82
DTIC

RESEARCH REPORT

PJH - 8

ROBUST PRINCIPAL COMPONENTS AND DISPERSION MATRICES VIA PROJECTION PURSUIT

Zhonglian Chen and Guoying Li

September 1981

This work was facilitated in part by National Science Foundation Grant MCS-79-08685 and Office of Naval Research Contract N00014-79-C-0512.

Department of Statistics
Harvard University
Cambridge

ROBUST PRINCIPAL COMPONENTS AND DISPERSION MATRICES VIA PROJECTION PURSUIT

Zhonglian Chen and Guoying Li*
Harvard University, Cambridge, MA

ABSTRACT

This paper discusses a new kind of robust procedure for estimating covariance/correlation matrices and their principal components. Robust eigenvectors and eigenvalues of a covariance matrix are obtained by the projection purusit method (PP) with robust variance as a projection index. Monte Carlo simulation results show that the best of the three projection pursuit type procedures introduced in this study compares favorably with approaches based on M-estimators of covariance: the estimate obtained by *the new procedure has about the same bias and variance as the best* M-estimators, and a *somewhat* better breakdown point.

KEYWORDS: Multidimensional data analysis, Robust estimation, Projection pursuit, Principal components, Covariance matrix, Correlation matrix.

*The authors are on leave from the Institute of Systems Science, Academia Sinica, Beijing, China.

I. INTRODUCTION

The classical covariance/correlation matrices and their eigenvalues and eigenvectors are widely used in multivariate data analysis. Unfortunately, they are sensitive to outliers--see Devlin et al. (1975)--and so are not robust.

The classical approach is: first, based on the data $\underset{\sim}{x}_1, \ldots, \underset{\sim}{x}_n$, the classical location estimate

$$\underset{\sim}{m} = \sum_{j=1}^{n} \underset{\sim}{x}_j / n \tag{1.1}$$

and the classical covariance matrix estimate

$$C = \sum_{j=1}^{n} (\underset{\sim}{x}_j - \underset{\sim}{m})(\underset{\sim}{x}_j - \underset{\sim}{m})'/n \tag{1.2}$$

are computed, and then the corresponding correlation matrix estimate R can be obtained by rescaling. Second, if necessary, the eigenvalues λ_i and eigenvectors $\underset{\sim}{a}_i$ $(i = 1, \ldots, p)$ of the matrix C (or R) can be computed by *using some* well *known* algorithm for solving the algebraic eigenproblem; thus

$$C = A \begin{bmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_p \end{bmatrix} A' \tag{1.3}$$

where $A = [\underset{\sim}{a}_1, \underset{\sim}{a}_2, \ldots, \underset{\sim}{a}_p]$ is an orthogonal matrix.

Along these lines a few robust estimators for covariance/correlation matrices have been proposed (cf. Huber (1981), Ch. 8). Elementwise approaches involve robust estimation of the individual elements of covariance/correlation matrices (Devlin et al. (1975, 1981)). Matrix approaches include the ellipsoidal multivariate trimming procedure (MVT), and maximum likelihood estimation of the shape matrix of some elliptical distribution (M-). Among the M-estimators, one is the maximum likelihood estimator for the t-distribution with f degrees of freedom (Maronna (1976)), usually f = 1 (MLT), another is the M-estimator with Huber-type weight (HUB) (Huber (1977)).

These robust methods have been studied either theoretically (Maronna (1976); Huber (1977); Huber (1981), Ch. 8) or via Monte Carlo (Devlin et al. (1981)). *M-estimators perform well on many criteria*, but unfortunately their breakdown points can be disappointingly low for high dimensional data. This means that although they are called robust, they may sometimes be influenced unduly by a few outliers in high dimensions.

This paper investigates an alternative kind of robust procedure for estimating principal components and covariance/correlation matrices. The procedure operates in reverse of the usual order of computations of covariances and eigenvectors: it first estimates robust eigenvectors and eigenvalues by a projection pursuit (PP) method (Friedman and Tukey (1974)), and then it constructs the estimated covariance matrix from these eigenvectors and eigenvalues according to (1.3). The method was proposed by Huber (1981, p.200) whose idea was "to mimic an eigenvector/eigenvalue determination and find the direction with the smallest or largest robust

variance, leading to an orthogonally invariant approach." *This approach* is conceptually simple, but it is not so easy to analyze theoretically, and its asymptotic variance etc. have not yet been obtained. We analyze, *as the first stage, performances and breakdown properties using the same* Monte Carlo techniques which are used by Devlin et al. (1981) to study several other robust dispersion matrix estimators. Some theoretical results will soon appear in a companion research report.

Our simulations show that the ACIA procedure, based on the average of the two covariance matrices estimated by the MIN PP procedure and the MAX PP procedure compares favorably with the best M-estimators HUB, MLT. *It seems to provide nearly as good performance as M-estimators but* with better breakdown properties. Consequently, this new kind of PP type procedure should be better known.

II. PROJECTION PURSUIT (PP) TYPE ESTIMATION OF PRINCIPAL COMPONENTS AND DISPERSION MATRICES

2.1 MIN PP Procedure and MAX PP Procedure

A sample of p dimensional data $\underset{\sim}{x}_1,\ldots,\underset{\sim}{x}_n$ may be projected onto a projection direction (unit vector) $\underset{\sim}{a}$, to obtain a one-dimensional sample of the projections $\{\underset{\sim}{a}'\underset{\sim}{x}_j\}$ of the data. As the projection direction vector $\underset{\sim}{a}$ varies over the unit sphere, different one-dimensional projected samples are obtained. If we take some variance estimate $v(\underset{\sim}{a})$ of the sample $\{\underset{\sim}{a}'\underset{\sim}{x}_j\}$ as a projection index, the MIN PP procedure is: p projection directions are determined sequentially so that subject to the constraint that $\underset{\sim}{a}_i$ be orthogonal to $\underset{\sim}{a}_{i-1},\ldots,\underset{\sim}{a}_1$, $v(\underset{\sim}{a})$ reaches its minimum value λ_i at $\underset{\sim}{a}_i$ $(i = 1,\ldots,p)$, that is

$$\left.\begin{aligned} \lambda_1 &= \min_{|\underset{\sim}{a}|=1} v(\underset{\sim}{a}) = v(\underset{\sim}{a}_1) \\ \lambda_2 &= \min_{\substack{|\underset{\sim}{a}|=1 \\ \underset{\sim}{a}\perp\underset{\sim}{a}_1}} v(\underset{\sim}{a}) = v(\underset{\sim}{a}_2) \\ &\;\vdots \\ \lambda_p &= \min_{\substack{|\underset{\sim}{a}|=1 \\ \underset{\sim}{a}\perp\underset{\sim}{a}_1,\ldots,\underset{\sim}{a}_{p-1}}} v(\underset{\sim}{a}) = v(\underset{\sim}{a}_p) \end{aligned}\right\} \tag{2.1}$$

To indicate the use of the MIN procedure in determining these quantities, we will label them $\lambda_i(\text{MIN})$, $\underset{\sim}{a}_i(\text{MIN})$.

Estimates from the MAX PP procedure are determined by:

$$\left.\begin{array}{l} \lambda_p = \max\limits_{|\underset{\sim}{a}|=1} v(\underset{\sim}{a}) = v(\underset{\sim}{a}_p) \\ \\ \lambda_{p-1} = \max\limits_{\substack{|\underset{\sim}{a}|=1 \\ \underset{\sim}{a} \perp \underset{\sim}{a}_p}} v(\underset{\sim}{a}) = v(\underset{\sim}{a}_{p-1}) \\ \vdots \end{array}\right\} \qquad (2.2)$$

and will be denoted by $\lambda_i(MAX)$, $\underset{\sim}{a}_i(MAX)$.

If the projection index is just the classical variance estimate

$$v(\underset{\sim}{a}) = \sum_{j=1}^{n} (z_i - \bar{z})^2/n \qquad (2.3)$$

where

$$z_j = \underset{\sim}{a}'\underset{\sim}{x}_j, \; j = 1,\ldots,n, \qquad (2.4)$$

it is well known that these two procedures provide the eigenvalues and eigenvectors of the classical covariance matrix estimate, so from (1.3), these two procedures provide the same covariance matrix estimate as the classical covariance matrix estimate (1.2) and

$$\bar{z} = \underset{\sim}{a}'\underset{\sim}{m} \qquad (2.5)$$

where $\underset{\sim}{m}$ is defined as (1.1).

If $v(a)$ is a robust estimate of the variance of the projected sample $\{\underset{\sim}{a}'\underset{\sim}{x}_j\}$, then these λ_i, $\underset{\sim}{a}_i$ can be considered as robust estimates of the eigenvalues and eigenvectors of a covariance matrix, and a robust estimate of the covariance matrix can be obtained according to (1.3). In general, the MIN PP procedure and the MAX PP procedure may give different results.

Via the Monte Carlo method, we intend to see: (i) whether these procedures work; (ii) their performances and breakdown properties; and (iii) which of the two procedures is better.

2.2 ACIA PP Procedure

In the process of developing the program, we observed that the eigenvalues of the covariance matrix estimated from the MIN PP procedure tended to be under the classical estimates and those estimated from the MAX PP procedure tended to be above the classical estimates at the normal distribution. We guessed that an average procedure, in some sense, of the MIN PP procedure (as the lower) and the MAX PP procedure (as the upper) would be closer to the classical estimator, which is very good for normal distribution, and should be better than these two at the normal distribution; this average procedure probably remains better than the other two at the contaminated normal distributions.

The ACIA PP procedure is based on the average of the two covariance matrices estimated by the MIN PP procedure and the MAX PP procedure, that is,

$$C(ACIA) = (C(MIN) + C(MAX))/2 \qquad (2.6)$$

and R(ACIA) is obtained by rescaling C(ACIA).

An analogical average procedure (ARIA) is based on

$$R(ARIA) = (R(MIN) + R(MAX))/2. \qquad (2.7)$$

Although C(ARIA) can be obtained by estimating the robust variance of each coordinate component, it loses the property of orthogonal invariance. So, we mainly compare the MIN, MAX, ACIA procedures in the study, though the results for the ARIA procedure are appended.

III. ALGORITHM

It requires some complicated implementation to make these procedures really work. We have made a time-consuming effort to design the algorithm and to program it carefully.

The algorithm for the MIN PP procedure used in this study consists of three hierarchies of iteration:

- outer: multivariate location - scatter iteration;
- middle: multivariate scatter step (MIN projection pursuit);
- inner: robust variance estimate on a projection direction.

3.1 Multivariate Location - Scatter Iteration

It is designed to be similar to the Huber type M-estimation (Huber (1981), p. 238) for making the new procedure have good performance, but the multivariate scatter step is replaced.

(1) Starting values: we take robust starting values

$$m_i = \operatorname{med}_j(x_{ij}), \quad i=1,\dots,p \tag{3.1}$$

$$A = I \tag{3.2}$$

$$\lambda_i = (\operatorname{med}_j(|x_{ij} - m_i|)/.6745)^2 \tag{3.3}$$

$$C = A \begin{bmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_p \end{bmatrix} A' \tag{3.4}$$

An alternative might be to use the classical estimates.

(2) First multivariate scatter step (see Section 3.2):

This follows by the iterations as follows.

(3) Multivariate location step:

$$C^{-1} := A \begin{bmatrix} 1/\lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & 1/\lambda_p \end{bmatrix} A' \tag{3.5}$$

(0 will be used instead of $1/\lambda_i$ when λ_i is too small, that is, C^{-1} is a generalized inverse); the square of the ellipsoidal distance

$$r_j^2 := (\underset{\sim}{x}_j - \underset{\sim}{m})' C^{-1} (\underset{\sim}{x}_j - \underset{\sim}{m}); \tag{3.6}$$

and the Huber-type weig

$$w_1(r_j) := \begin{cases} 1 & r_j \le \kappa_1 \\ \kappa_1/r_j & r_j > \kappa_1 \end{cases} \qquad i=1,\dots,n \tag{3.7}$$

are evaluated. The squared corner constant κ_1^2, as a tunable parameter is always taken as the 90%-point of χ_p^2 distribution in this study. Then the multivariate location is updated by

$$h := \sum w_1(r_j)\,(x_j - m)/\sum w_1(r_j) \tag{3.8}$$

$$m := m + h \tag{3.9}$$

(4) Multivariate scatter step (see section 3.2): The current λ_i, A, C are evaluated.

(5) Termination rule: Suppose $X \sim N(0,C)$. Let $Y = U(X - m)$

$$U = \begin{bmatrix} \lambda_1^{-1/2} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_p^{-1/2} \end{bmatrix} A'; \tag{3.10}$$

then $C^{-1} = U'U$ and $Y \sim N(0,I)$. So, we define

$$e_l = \|Unew - Uold\| / \|Uold\| \tag{3.11}$$

where

$$\|Uold\|^2 = tr(Uold * Uold')$$

$$= \sum_{i=1}^{p} 1/\lambda old_i \tag{3.12}$$

$$\|Unew - Uold\|^2 = tr((Unew - Uold)'(Unew - Uold))$$

$$= \sum_{i,k=1}^{p} (anew(i,k)/\lambda new_k^{1/2} - aold(i,k)/\lambda old_k^{1/2}) \tag{3.13}$$

and define

$$f_l^2 = \|Unew * h\|^2$$

$$= \sum_{k=1}^{p} \left(\sum_{i=1}^{p} anew(i,k) * h_i \right)^2 / \lambda new_k \tag{3.14}$$

The iterations are terminated when either $e_l < \epsilon_1$ and $f_l \le \delta_1$ or the iteration number $i_l \ge M_1$. The tunable parameters are taken as $\epsilon_1 = 10^{-4}$, $\delta = \infty$ (the multivariate location is considered as a *nuisance*) *and* $M_1 = 15$ here.

3.2 <u>The Multivariate Scatter Step</u>

λ_i, $a_i (i = 1,\dots,p)$, A, C are updated in this step. Suppose a robust estimate $v(a)$ of the variance of the projections of the data *onto any direction* a is available (see Section 3.3). The following process is repeated for $i = 1,\dots,p$.

<u>3.2.1</u>

Suppose that the latest i-1 optimal projection directions $a_1,\dots,a_{i-1}$ being mutually orthogonal, the corresponding $\lambda_1,\dots,\lambda_{i-1}$ and the orthogonal matrix

$$A: = [a_1,\dots,a_{i-1} \vdots a_i,\dots,a_p] \tag{3.15}$$

are available. $a_i,\dots,a_p$ *are an orthonormal* basis of the orthogonal complement subspace of $a_1,\dots,a_{i-1}$. Suppose $a \perp a_1,\dots,a_{i-1}$, then

$$a = [a_i,\dots,a_p]\alpha = \sum_{k=1}^{p-i+1} \alpha_k a_{i-1+k} \tag{3.16}$$

where $\alpha = [\alpha_1,\dots,\alpha_{p-i+1}]' \in R^{p-i+1}$

The next optimal projection a_i, is searched for such that

$$\lambda_i = \min_{|a|=1} v(a) = v(a_i)$$

$$= \min_{|\alpha|=1} v(\alpha) = v(\alpha_i). \tag{3.17}$$

It is equivalent to finding an optimal direction α_i with minimal projection index $v(\alpha_i)$ in p-i+1 dimensional space.

(1) Starting direction.

The starting direction is given by choosing an optimal direction from the p-i+1 coordinate axes in the p-i+1 dimensional space. An alternative may be to generate a random starting direction.

(2) The MIN subroutine in p-i+1 dimensional space.

The algorithm drawn from the PPR program (Friedman et al. (1979)) is as follows. An orthonormal coordinate system in which the given starting direction is its first axis, is generated from the given *starting direction by using an elementary orthogonal transformation.* An optimal direction is found by searching along the unit circle in the two dimensional plane spanned by the 1-2 axes and using a parabolic interpolation refinement; a further optimal direction is found in the next plane spanned by the former optimal direction and the next axis, ..., and so on, until completing a step of iteration. A new orthonormal coordinate system is generated from the latest optimal direction, ..., and so forth. The iterations continue until $|vnew^{1/2} - vold^{1/2}|/vold^{1/2} \le \varepsilon_2$ or the iteration number $i_2 \ge M_2$. The tunable parameters are taken as $\varepsilon_2 = 10^{-4}$, $M_2 = 15$ here. At this point, the ith optimal projection direction

$$\underset{\sim}{a}_i = [\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_p]\underset{\sim}{\alpha}_i \tag{3.18}$$

and the corresponding $\lambda_i = v(\underset{\sim}{a}_i)$ are available.

3.2.2

That vector *among the old* $\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_p$ *which has the largest* projection length onto the new $\underset{\sim}{a}_i$ is replaced by the old $\underset{\sim}{a}_i$; then by using Gram-Schmidt orthogonalization process, the new $\underset{\sim}{a}_{i+1}, \ldots, \underset{\sim}{a}_p$ are obtained and the orthogonal matrix is updated using

$$A: = [\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_i \vdots \underset{\sim}{a}_{i+1}, \ldots, \underset{\sim}{a}_p], \tag{3.19}$$

where $\underset{\sim}{a}_1, \ldots, \underset{\sim}{a}_i$ are the latest projection directions being mutually orthogonal.

For $i = 1, \ldots, p$, all λ_i, $\underset{\sim}{a}_i$ are evaluated out, then the latest covariance matrix C is obtained.

3.3 Robust Variance Procedure on a Projection Direction

The location for the one dimensional sample

$$z_j = \underset{\sim}{a}'\underset{\sim}{x}_j, \quad j = 1, \ldots, n \tag{3.20}$$

is evaluated from the current multivariate location vector m, by

$$z^*: = \underset{\sim}{a}'\underset{\sim}{m} \tag{3.21}$$

and it is shifted out of z,

$$d_j = z_j - z^*. \tag{3.22}$$

Then the M-estimate v(a) is evaluated based on $\{d_j\}$ as follows.

(1) Initial value for variance is given from the current covariance matrix, by

$$v: = \underset{\sim}{a}'C\underset{\sim}{a} \tag{3.23}$$

(2) The Huber-type weights are given by

$$w_3(d_j^2) := \begin{cases} 1 & d_j^2 \le \kappa_3^2 v \\ v\kappa_3^2/d_j^2 & d_j^2 > \kappa_3^2 v \end{cases} \qquad (3.24)$$

The variance is updated according to

$$v := \left(\sum_{j=1}^{n} w_3(d_j^2)\, d_j^2/n \right) / B(\kappa_3) \qquad (3.25)$$

where $B(\kappa_3)$ is the factor which makes the variance estimator be consistent under normal distribution (Huber (1981)).

$$B(\kappa_3) = \int_{|x|<\kappa_3} x^2\, d\Phi(x) + 2\kappa_3^2 (1 - \Phi(\kappa_3))$$

$$= G(\kappa_3^2/2; 1.5) + 2\kappa_3^2 (1 - \Phi(\kappa_3)); \qquad (3.26)$$

here $\Phi(x)$ is the standard normal distribution function and $G(x,f)$ is the Gamma distribution function with f degrees of freedom. The tunable parameter κ_3 is taken as 1.5 in this study.

(3) The iteration termination rule is defined by

$$|vnew^{1/2} - vold^{1/2}| / vold^{1/2} \le \epsilon_3$$

or the iteration number $i_3 \ge M_3$. The tunable parameters are taken as $\epsilon_3 = 10^{-4}$, $M_3 = 20$ here.

In this study, the location z* is fixed for all these iterations. An alternative might be that z* from (3.21) is only used as an initial value and one dimensional location-scale M-estimators z*, v are computed.

Other robust variance estimates can be used, for example, the square of the median absolute deviation as in (3.3).

As soon as eigenvectors and eigenvalues are obtained, the covariance matrix C is given by (1.3).

For ACIA PP procedure, we run the MIN PP procedure first, then run the MAX PP procedure independently, then the covariance matrix is given by (2.6), though it might be time-saving to run the MAX procedure first and then to run the MIN procedure using the results from the former as the starting values.

IV. SIMULATION CONDITIONS

The Monte Carlo simulations which we employed in studying the MIN, MAX and ACIA projection pursuit procedures was patterned after that used by Devlin et al. (1981). *In our experiment, we evaluated performances* only for the three new procedures and the classical approach (which was used as a control group). Devlin et al.'s study examined a variety of estimators under several different conditions. By keeping our simulation conditions parallel to theirs and using their results, we can indirectly compare the behavior of the new procedures with the behavior of the methods which they have already studied.

4.1 For Testing Performance

Dimension $p = 6$

$$\text{Target:} \quad C = P = \begin{bmatrix} P_1 & 0 \\ 0 & P_2 \end{bmatrix} = A \begin{bmatrix} \lambda_1 & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_6 \end{bmatrix} A' \tag{4.1}$$

with

$$P_1 = \begin{bmatrix} 1 & & \\ .95 & 1 & \\ .30 & .10 & 1 \end{bmatrix} \tag{4.2}$$

$$P_2 = \begin{bmatrix} 1 & & \\ -.499 & 1 & \\ -.499 & -.499 & 1 \end{bmatrix} \tag{4.3}$$

$$\left.\begin{array}{lll} \lambda_1 = .002 & \lambda_2 = .02849166 & \lambda_3 = .9429570 \\ \lambda_4 = \lambda_5 = 1.499000 & & \lambda_6 = 2.028551 \end{array}\right\} \tag{4.4}$$

$$A = \begin{bmatrix} 0 & .7145431 & -.8243439 & 0 & 0 & .6947176 \\ 0 & -.6832509 & -.2956021 & 0 & 0 & .6676734 \\ 0 & -.1503027 & -.9517478 & 0 & 0 & .2575438 \\ -.5773503 & 0 & 0 & -.4082483 & .7071068 & 0 \\ -.5773503 & 0 & 0 & -.4082483 & -.7071068 & 0 \\ -.5773503 & 0 & 0 & .8164966 & 0 & 0 \end{bmatrix} \tag{4.5}$$

Sample size $n = 50$, replication number $m = 200$.

$$\text{Let } V = A \begin{bmatrix} \lambda_1^{1/2} & & & \\ & \cdot & & \\ & & \cdot & \\ & & & \lambda_p^{1/2} \end{bmatrix} \tag{4.6}$$

We use four simulation sample models:

(1) Normal: NOR[0,P]

$$x_j = Vy_j, \qquad j = 1,\ldots,n \tag{4.7}$$

(2) Symmetric Contaminated Normal: SCN[0,P]

$$x_j = b_j V\mathbf{y}_j + 3(1-b_j)V\mathbf{y}_j \qquad (4.8)$$

(3) Cauchy: CAU[0,P]

$$x_j = V\mathbf{y}_j/\omega_j \qquad (4.9)$$

(4) Asymmetric Contaminated Normal: ACN[μ,P]

$$x_j = b_j V\mathbf{y}_j + (1-b_j)(V\mathbf{y}_j + \mu) \qquad (4.10)$$

where $\mathbf{y}_j$ is the sample from NOR[0,I], generated by Box-Muller technique, b_j is the sample from the independent Bernoulli distribution with $Pr(B=0) = 0.1$ and ω_j is the sample from the standard half-normal distribution independently, $\mu = -0.537 * a = [0, 0, 0, .310037, .310037, .310037]'$

4.2 For Testing Empirical Breakdown Point

Dimension $p = 20$

Sample Size $n = 1000$

Replication $m = 1$

(1) Symmetric contamination. A NOR[0,I] is subjected to the α% symmetric contamination for which

$$\left.\begin{array}{ll} x_j = 100\,\mathbf{y}_j, & j = 1,\ldots,10\alpha \\ x_j = \mathbf{y}_j, & j = 10\alpha+1,\ldots,1000 \end{array}\right\} \qquad (4.11)$$

(2) Asymmetric contamination. A NOR[0,I] is subjected to the α% asymmetric contamination for which

$$\left.\begin{array}{ll} x_j = 100\,\mathbf{1}, & j = 1,\ldots,10\alpha \\ x_j = \mathbf{y}_j & j = 10\alpha+1,\ldots,1000 \end{array}\right\} \qquad (4.12)$$

where $\mathbf{y}_j$ are from NOR[0,I].

V. SIMULATION RESULTS

The three new PP type procedures MIN, MAX and ACIA are compared with the classical estimator based on the results of the simulations described in Section IV and are indirectly compared with the other robust estimators based on the simulation results of Devlin et al. (1981, section 4).

5.1 Correlation Coefficient Comparisons

The average bias (times 1000) of the *individual correlation* coefficients, and the mean square error (MSE) (times 1000) of the Z-transformation $z = 0.5 \log[(1+\rho)/(1-\rho)]$ are listed in Table 1 for each element of P_1 and P_2 which is neither 0 or 1 and for the worst case among the nine estimates of zero.

A. MSE

ACIA, based on the average of the covariance matrices estimated by the MIN PP procedure and the MAX PP procedure, generally has the smallest MSE of the PP procedures studied in this paper.

Comparing Table 1 in this paper with the one given in Devlin et al. (1981), it appears that

- That classical R exhibits the same nonrobust behavior in our simulations as was evident in those of Devlin et al. (1981). The MSE's for the classical R in this study agree closely with those in Devlin et al. (1981). This indicates that our simulation results should be comparable with theirs. Large differences between the two simulation studies are evident only at the Cauchy distribution, where the Monte Carlo sampling variance is large anyway, and such differences might be anticipated.
- Across all models, the MSE's of ACIA in this study lie between *those of the HUB and MLT M-estimators* computed by Devlin et al. (1981) (there is an exception in the Cauchy simulation).

B. Bias

The MIN and MAX methods have biases of roughly the same sizes. The biases of the ACIA method, as if averages of those of the MIN and *MAX procedures, do somewhat better in each case* than whichever of the MIN and MAX procedures is worst.

A comparison of the bias obtained here with those in Table 1 of Devlin et al. may be unreliable. The biases of the classical estimator of R show discrepancies between the two tables. Keeping this in mind, the biases for the PP procedures seem larger than the biases found in Devlin et al. for the HUB and MLT M-estimators.

5.2 Eigenvalue and Eigenvector Comparisons for Correlation Matrix

The average bias of eigenvalues (times 1000) and the MSE of log eigenvalues (times 100) for λ_i, $i = 1,\ldots,6$ is recorded in Table 2.

A. Eigenvalues

A glance at Table 2 will show that the ACIA procedure performs best among the PP procedures. Comparing Table 2 in this paper with the one

given in Devlin *et al.* (1981), we see that

- *The lack of robustness of the classical eigenvalue estimator is apparent in both of Devlin et al.'s and our simulation study.* The simulation results of Devlin et al. for the classical eigenvalue estimator agree closely with the results in this study.
- Comparing across studies, the ACIA procedure performs nearly as well as the M-estimators; in particular, ACIA appears quite competitive with HUB.

B. Eigenvectors

The ACIA procedure also gives good estimates of eigenvectors. *Table 3* displays the mean of $|\cos\theta_i|$ (*note: here the* λ_i *is ordered as* $\lambda_1 > \lambda_2 = \lambda_3 > \ldots > \lambda_6$). For $i \neq 2,3$, θ_i *is the angle between the ith target eigenvector and its estimate; for* $i = 2$ *or* 3, θ_i *is the angle between the ith eigenvector estimate and the subspace spanned by the second and third target eigenvectors. The ACIA procedure does fairly well for estimating eigenvectors corresponding to eigenvalues of all sizes.*

5.3 Empirical Breakdown

The breakdown point (Hampel (1971)) gives the maximum fraction of bad outliers which the estimator can cope with. Devlin et al. defined the empirical breakdown point for estimators of a 20×20 identity matrix as follows: "Breakdown was judged to have occurred if the average absolute correlation was as large as 0.05 or $\hat{\lambda}_1/\hat{\lambda}_{20}$ for the correlation matrix was 3.0 or larger" (Devlin et al. (1981), section 4.5).

We desired to explore breakdown properties and performances with high accuracy computations. Very high computational precisions $\epsilon_1 = \epsilon_2 = \epsilon_3 = 10^{-4}$ for terminating iterations were specified. The maximal iteration numbers were chosen as $M_1 = 15$, $M_2 = 15$, and $M_3 = 20$. In the simulations for $p = 6$, the specified precisions were almost always achieved within the allowed number of iterations. In the breakdown study, where now $p = 20$, we kept the same tolerances and iteration limits, and we believe that with these specifications, the computations of empirical breakdown are accurate. For example, a typical application of the MAX procedure in the breakdown study had $i_1 = 15$, $e_1 = .002$, and $f_1 = .0005$.

The results for the MAX, MIN and ACIA procedures indicate that empirical breakdown does not occur for symmetric contamination of even 25%. (Its high theoretical breakdown point will be discussed in our next report.) In the case of asymmetric contamination, *the estimators* do not breakdown at 5% contamination; but *at the 10% level, the MIN and* ACIA procedures begin to show breakdown. *The cause of breakdown seems to be the large bias of multivariate location which makes the robust variance estimate inexact (cf. Section III and the discussion in section* 3.3).

Comparing now with the results in Devlin et al. (1981) (section 4.5 Table 3), the empirical breakdown point of the projection pursuit type procedures as measured in this study appears to be higher than the M-estimators studies in Devlin et al. (1981) and slightly better than trimming procedure MVT (with robust start).

5.4 Computation Time Comparison

A VAX-11 DEC (Digital Equipment Corporation) computer was used to perform the computations in this study. The central processor unit (CPU) time required by the VAX for a problem of size $p = 6$, $n = 50$ is shown in Table 4. On this VAX, 15 seconds of prime CPU time costs about 30 cents, so for small problems, the cost of the PP approach is quite reasonable.

Table 4 shows that the MIN procedure needs more CPU time than the MAX procedure and the computation time grows with the heaviness of the tail and the asymmetry of a distribution.

In the empirical breakdown study, where the problem size was $p = 20$, $n = 1000$, the ACIA procedure took about 5 CPU hours, which at the cheapest rate on our VAX would cost about $40. (Of course, this was an artificial high asymmetric problem.) It seems that the computation time is proportional to np^3 for this computation.

5.5 Results for Covariance Matrix Under Normal

Table 5 lists the bias and MSE of the covariance elements estimator under the normal distribution. Table 6 reports the bias of the eigenvalues and the MSE of log eigenvalues. The mean of $|\cos\theta|$ for covariance matrix is listed in Table 7. The performances of ACIA are really much closer to those of the classical estimators (which are optimal at the normal distribution) than the other two.

VI. SUMMARY

The Projection Pursuit type procedures MIN, MAX and ACIA are a new kind of robust procedure for principal components and dispersion matrices. They are orthogonally invariant, and by design should have good breakdown and efficiency robustness. They are competitors with the best robust procedures based on M-estimators of the covariance matrix.

This study describes an implementation of the PP procedures and used Monte Carlo simulations patterned after those of Devlin et al. (1981) to evaluate the performances and breakdown properties of PP procedures (mainly for estimates of the correlation matrix and its eigenvalues and eigenvectors, although properties of estimates of covariances were studied as well). The parallelism of our study with that of Devlin et al. allows a comparison of our results for PP with the ones they evaluated for other robust procedures.

In general, the MIN and MAX projection pursuit procedures give different results. This study proposes an idea that an average procedure, in some sense, might be better than both of the MIN and MAX procedures and provides an average procedure ACIA. This idea may be helpful in a projection pursuit setting involving other kinds of projection indexes.

The ACIA procedure is the best overall of the three procedures. The ACIA procedure

- estimates eigenvalues and eigenvectors of the correlation matrix as well as the M-estimators studies by Devlin et al.
- estimates individual elements of the correlation matrix nearly as well as the M-estimators with MLT or HUB weights, when the mean square error of the Z-transformed elements is considered.

The breakdown point of these three new procedures seems to be higher than those of the M-estimators.

A potential shortcoming of this implementation for the PP procedures is the computational expense involved in its use with high dimensional data. We have not yet observed to what extent the tunable parameters of the algorithm can be adjusted so as to reduce the amount of computation required without adversely affecting the efficiency or breakdown properties of the method.

Future avenues for research might include:

(1) Testing other robust variance estinators to see if breakdown or efficiency properties may be improved. The median absolute deviation scale estimator is particularly resistant to breakdown and might perform well in some contexts.

(2) Designing a fast algorithm for high dimensional data.

(3) Looking for other average procedures, in some sense, of the MIN and MAX procedures.

We will explore the theoretical properties of PP estimators, including such issues as consistency and breakdown point, in a forthcoming research report.

VII. ACKNOWLEDGEMENTS

The authors are grateful to Professor Peter J. Huber for his guidence to this topic and for his enlightening discussion and comments. We are thankful to David Donoho for his help in revising the manuscript and constructive suggestions.

REFERENCES

Andrews, D.F. et al. (1972), Robust Estimates of Location: Survey and Advances, Princeton University Press, Princeton, N.J.

Chen, H., Gnanadesikan, R., Kettenring, J.R. (1974), "Statistical Methods for Grouping Corporations," Sankhya Ser. B 36, 1-28.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1975), "Robust Estimation and Outlier Detection with Correlation Coefficients," Biometrika 62, 531-545.

Devlin, S.J., Gnanadesikan, R., and Kettenring, J.R. (1981), "Robust Estimation of Dispersion Matrices and Principal Components," Journal of the American Statistical Association 76, 354-362.

Friedman, J.H., Jacobson, M., Stuetzle, W. (1980), "Projection Pursuit Regression," Tech. Rep. No. 146, Dept. of Statistics, Stanford Univ.

Friedman, J.H., Stuetzle, W., Chaffee, R.C. (1979), "PPR Program," Stanford Linear Accelerator Center.

Friedman, J.H., Tukey, J.W. (1974), "A Projection Pursuit Algorithm for Exploratory Data Analysis," IEEE Transactions on Computers, Vol. C-23, 9, 881-890.

Gnanadesikan, R., and Kettenring, J.R. (1972), "Robust Estimates, Residuals, and Outlier Detection with Multiresponse Data," Biometrics 28, 81-124.

Hampel, F.R. (1971), "A General Qualitative Definition of Robustness," Annals of Mathematical Statistics 42, 1887-1896.

Huber, P.J. (1964), "Robust Estimation of a Location Parameter," Annals of Mathematical Statistics 35, 73-101.

Huber, P.J. (1977), "Robust Covariances," in Statistical Decision Theory and Related Topics II, ed. Gupta, S. and Moore, D.S., Academic Press, New York, 165-191.

Huber, P.J. (1981), Robust Statistics, John Wiley & Sons, New York.

Maronna, R.A. (1976), "Robust M-Estimators of Multivariate Location and Scatter," Annals of Statistics 4, 51-67.

Mosteller, F., and Tukey, J.W. (1977), Data Analysis and Regression, Addison-Wesley, Reading, MA.

Table 1. 1,000 × Bias of Correlation Estimators and 1,000 × MSE of Their z-Transforms

		1,000 × BIAS					1,000 × MSE				
DIST'N	1,000 × ρ	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)
	950	-1	-6	-4	-5	-5	20	34	24	24	25
	300	-4	-39	-17	-27	-28	19	38	24	25	25
	100	-3	-45	-21	-32	-33	19	39	25	27	27
NOR	-499	1	-2	-7	-6	-4	21	32	25	24	25
	-499	-8	2	-3	-1	0	21	29	25	24	24
	-499	19	16	24	20	20	22	29	26	24	24
	0-max	17	19	18	19	19	24	35	27	26	26
	950	-3	-3	-7	-5	-5	49	34	35	30	29
	300	-12	-25	-17	-20	-21	59	39	31	29	29
	100	-7	-20	-24	-23	-22	54	38	32	29	30
SCN	-499	3	12	16	14	14	57	39	28	28	29
	-499	14	-9	-11	-11	-10	55	31	27	24	25
	-499	10	15	10	11	13	46	29	25	23	23
	0-max	-22	-23	-23	-22	-22	59	41	30	28	29
	950	-49	-3	-36	-22	-19	869	53	126	77	67
	300	-108	-54	-40	-44	-47	976	66	52	44	44
	100	-101	-42	-61	-53	-52	931	62	52	41	41
CAU	-499	68	42	25	30	34	1083	65	43	42	44
	-499	86	-1	7	3	3	1131	44	49	37	37
	-499	44	-2	-4	-5	-3	1180	58	44	40	41
	0-max	66	17	33	19	18	1114	50	45	33	33
	950	-1	-4	-6	-5	-5	22	36	29	29	29
	300	-3	-13	-18	-14	-15	20	40	29	29	29
	100	1	-11	-20	-15	-15	20	38	29	28	28
ACN	-499	18	13	18	15	16	23	35	27	26	27
	-499	13	4	0	1	2	21	32	24	23	24
	-499	20	7	20	14	14	19	30	24	23	23
	0-max	15	15	-12	13	13	24	37	31	28	29

Table 2. 1,000 × Bias of Eigenvalue Estimators and 100 × MSE of Their Logs

		1,000 × BIAS						100 × MSE					
DIST'N	λ	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)	λ (AERIA)	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)	λ (AERIA)
	2.029	192	250	201	192	194	226	1	2	1	1	1	2
	1.499	110	119	121	111	111	120	1	1	1	1	1	1
NOR	1.499	-187	-223	-193	-189	-191	-208	3	4	3	3	3	3
	.943	-113	-146	-128	-115	-115	-137	3	6	4	4	4	5
	.028	-1	0	-1	-1	1	0	9	12	10	10	10	10
	.002	0	0	0	0	1	0	9	10	11	10	29	10
	2.029	391	292	220	216	220	256	4	2	2	2	2	2
	1.499	136	113	121	105	104	117	2	1	1	1	1	1
SCN	1.499	-302	-256	-207	-208	-212	-232	7	5	4	4	4	4
	.943	-223	-148	-133	-114	-114	-141	13	6	5	4	4	5
	.028	-3	-1	0	1	1	0	21	13	12	12	12	12
	.002	0	0	0	0	2	0	20	10	10	9	41	9
	2.029	1716	321	216	205	210	268	40	3	2	2	2	2
	1.499	-119	133	165	139	134	152	67	2	2	1	1	2
CAU	1.499	-887	-275	-235	-227	-232	-255	344	6	5	4	4	5
	.943	-629	-185	-159	-129	-128	-172	693	12	9	6	6	8
	.028	-17	1	12	12	12	7	658	25	31	28	28	21
	.002	-1	0	1	2	4	1	933	21	33	42	117	21
	2.029	202	285	217	216	219	251	1	2	2	2	2	2
	1.499	93	112	112	103	103	112	1	1	1	1	1	1
ACN	1.499	-205	-245	-212	-210	-213	-228	3	5	4	3	3	4
	.943	-114	-157	-131	-120	-121	-144	3	7	5	4	4	5
	.028	5	0	3	2	2	2	7	12	10	10	10	9
	.002	18	3	11	9	9	7	523	93	329	268	284	218

Note: λ(AERIA) defined as

$$\lambda(AERIA) = (\lambda(MIN) + \lambda(MAX))/2$$

where λ(MIN), λ(MAX) are eigenvalues of R(MIN), R(MAX).

Table 3. Mean of $|\cos \theta_i|$ for Correlation Matrix

DIST'N	λ	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)
			MEAN of $\|\cos \theta\|$			
NOR	2.029	.885	.849	.877	.876	.874
	1.499	.922	.896	.912	.913	.912
	1.499	.987	.872	.883	.891	.889
	.943	.914	.870	.887	.894	.892
	.028	.998	.996	.996	.997	.997
	.002	.996	.995	.995	.996	.995
SCN	2.029	.798	.845	.862	.872	.871
	1.499	.858	.904	.914	.927	.927
	1.499	.850	.858	.869	.878	.877
	.943	.852	.869	.884	.900	.899
	.028	.995	.995	.995	.996	.996
	.002	.992	.994	.994	.995	.994
CAU	2.029	.598	.816	.823	.836	.836
	1.499	.711	.877	.879	.888	.887
	1.499	.674	.838	.864	.874	.873
	.943	.585	.830	.852	.872	.871
	.028	.940	.983	.977	.981	.979
	.002	.948	.980	.982	.981	.978
ACN	2.029	.885	.846	.869	.872	.871
	1.499	.924	.904	.912	.918	.917
	1.499	.895	.863	.878	.887	.886
	.943	.911	.878	.882	.899	.899
	.028	.627	.954	.873	.931	.932
	.002	.625	.952	.372	.930	.930

Table 4. CPU Time on VAX-11 Computer for a Sample

DIST'N	p	n	MIN	MAX	ACIA
NOR	6	50	15 seconds	10 seconds	25 seconds
SCN	6	50	16	12	28
CAU	6	50	22	20	42
ACN	6	50	20	11	31
SCN (25%)	20	1000			5 hours
ACN (5%)	20	1000			4 hours
ACN (10%)	20	1000			5 hours
SCN (25%)	6	50		25 seconds	
ACN (25%)	6	50		43 seconds	
SCN (10%)	10	50			130 seconds

Table 5. 1,000 × Bias of Covariance Estimators and 1,000 × MSE of them for NOR

	1,000 BIAS				1,000 MSE			
ρ	C	C (MIN)	C (MAX)	C (ACIA)	C	C (MIN)	C (MAX)	C (ACIA)
1.	-14	-92	43	-24	39	65	52	49
1.	-10	-84	57	-13	39	61	55	48
1.	-19	-35	146	56	38	67	88	62
1.	-35	-133	57	-38	37	55	56	41
1.	-15	-108	94	-7	42	61	60	44
1.	-45	-122	63	-29	40	59	56	44
.95	-12	-87	44	-22	37	60	49	46
.30	-5	-52	12	-20	19	35	28	26
.10	-4	-47	-14	-30	18	33	28	26
-.499	34	75	-11	32	24	30	35	27
-.499	11	48	-51	-2	25	34	34	28
-.499	2	59	-45	7	23	28	34	25
.0	3	5	-1	2	19	23	24	20
.0	9	10	5	8	20	25	26	22
.0	-5	-5	-4	-5	16	25	25	21
.0	-6	-2	-9	-5	19	25	26	22
.0	-14	-8	-15	-12	21	26	28	23
.0	18	18	21	19	21	24	30	23
.0	2	-3	9	3	21	22	27	20
.0	5	-2	9	4	22	23	26	20
.0	-13	-11	-16	-14	21	28	31	25

Table 6. 1,000 × Bias of Eigenvalue Estimators and 100 × MSE of their Log for NOR

	1,000 × BIAS				100 × MSE			
λ	C	C (MIN)	C (MAX)	C (ACIA)	C	C (MIN)	C (MAX)	C (ACIA)
2.029	247	167	449	261	3	4	6	4
1.499	107	-309	285	121	2	3	5	3
1.499	-309	-449	-185	-297	8	17	5	8
.943	-164	-256	-87	-140	8	17	7	8
.028	-2	-4	0	-1	6	10	7	8
.002	0	0	0	0	7	10	7	7

Table 7. Mean of $|\cos \theta_i|$ for Covariance Matrix

DIST'N	λ	C	MEAN of $\|\cos \theta\|$ C (MIN)	C (MAX)	C (ACIA)
NOR	2.029	.780	.730	.711	.743
	1.499	.841	.797	.776	.798
	1.499	.812	.780	.763	.793
	.943	.818	.725	.735	.758
	.028	.998	.997	.997	.997
	.002	.999	.998	.998	.999
SCN	2.029	.634	.731	.703	.737
	1.499	.733	.796	.778	.800
	1.499	.767	.774	.737	.779
	.943	.719	.733	.710	.748
	.028	.996	.996	.995	.997
	.002	.998	.998	.998	.998
CAU	2.029	.463	.645	.609	.653
	1.499	.676	.740	.701	.741
	1.499	.654	.740	.705	.736
	.943	.569	.697	.607	.680
	.028	.970	.984	.978	.981
	.002	.987	.987	.988	.986
ACN	2.029	.780	.725	.704	.737
	1.499	.827	.810	.777	.804
	1.499	.826	.774	.768	.788
	.943	.815	.752	.735	.765
	.028	.624	.960	.883	.950
	.002	.624	.961	.883	.951

Table 1. 1,000 × Bias of Correlation Estimators and 1,000 × MSE of Their z-Transforms

		1,000 × BIAS					1,000 × MSE				
DIST'N	1,000 × p	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)	R	R (MIN)	R (MAX)	R (ACIA)	R (ARIA)
NOR	950	-1	-6	-4	-5	-5	20	34	24	24	25
	300	-4	-39	-17	-27	-28	19	38	24	25	25
	100	-3	-45	-21	-32	-33	19	39	25	27	27
	-499	1	-2	-7	-6	-4	21	32	25	24	25
	-499	-8	2	-3	-1	0	21	29	25	24	24
	-499	19	16	24	20	20	22	29	26	24	24
	0-max	17	19	18	19	19	24	35	27	26	26
SCN	950	-3	-3	-7	-5	-5	49	34	35	30	29
	300	-12	-25	-17	-20	-21	59	39	31	29	29
	100	-7	-20	-24	-23	-22	54	38	32	29	30
	-499	3	12	16	14	14	57	39	28	28	29
	-499	14	-9	-11	-11	-10	55	31	27	24	25
	-499	10	15	10	11	13	46	29	25	23	23
	0-max	-22	-23	-23	-22	-22	59	41	30	28	29
CAU	950	-49	-3	-36	-22	-19	869	53	126	77	67
	300	-108	-54	-40	-44	-47	976	66	52	44	44
	100	-101	-42	-61	-53	-52	931	62	52	41	41
	-499	68	42	25	30	34	1083	65	43	42	44
	-499	86	-1	7	3	3	1131	44	49	37	37
	-499	44	-2	-4	-5	-3	1180	58	44	40	41
	0-max	66	17	33	19	18	1114	50	45	33	33
ACN	950	-1	-4	-6	-5	-5	22	36	29	29	29
	300	-3	-13	-18	-14	-15	20	40	29	29	29
	100	1	-11	-20	-15	-15	20	38	29	28	28
	-499	18	13	18	15	16	23	35	27	26	27
	-499	13	4	0	1	2	21	32	24	23	24
	-499	20	7	20	14	14	19	30	24	23	23
	0-max	15	15	-12	13	13	24	37	31	28	29

From: Devlin, S. J., Gnanadesikan, R., and Kettenring, J. R. (1981), "Robust Estimation of Dispersion Matrices and Principal Components," Journal of the American Statistical Association, 76, 354-362.

DAT
FILM